

# Modeling of surface dust concentration in snow cover at industrial area using neural networks and kriging

Cite as: AIP Conference Proceedings **1836**, 020033 (2017); <https://doi.org/10.1063/1.4981973>  
Published Online: 05 June 2017

A. P. Sergeev, D. A. Tarasov, A. G. Buevich, A. V. Shichkin, A. G. Tyagunov, and A. N. Medvedev



View Online



Export Citation

## ARTICLES YOU MAY BE INTERESTED IN

[Performance evaluation of firefly algorithm with variation in sorting for non-linear benchmark problems](#)

AIP Conference Proceedings **1836**, 020032 (2017); <https://doi.org/10.1063/1.4981972>

[Topsoil pollution forecasting using artificial neural networks on the example of the abnormally distributed heavy metal at Russian subarctic](#)

AIP Conference Proceedings **1836**, 020024 (2017); <https://doi.org/10.1063/1.4981964>

[High variation subarctic topsoil pollutant concentration prediction using neural network residual kriging](#)

AIP Conference Proceedings **1836**, 020023 (2017); <https://doi.org/10.1063/1.4981963>



**SHFQA**  
Quantum Analyzer  
8.5 GHz

**Zurich Instruments**

**New**

## Your Qubits. Measured.

Meet the next generation of quantum analyzers

- Readout for up to 64 qubits
- Operation at up to 8.5 GHz, mixer-calibration-free
- Signal optimization with minimal latency

[Find out more](#)

**Zurich Instruments**

# Modeling Of Surface Dust Concentration In Snow Cover At Industrial Area Using Neural Networks And Kriging

Sergeev A.P.<sup>1,2,a)</sup>, Tarasov D.A.<sup>1,2,b)</sup>, Buevich A.G.<sup>2,c)</sup>, Shichkin A.V.<sup>1,2,d)</sup>,  
Tyagunov A.G.<sup>1,e)</sup>, and Medvedev A.N.<sup>1,2,f)</sup>

<sup>1</sup>*Ural Federal University, Mira, 19, Ekaterinburg, RUSSIA 620002*

<sup>2</sup>*Institute of Industrial Ecology UB RAS, Sophy Kovalevskoy, 20, Ekaterinburg, RUSSIA 620990*

*a) Corresponding author: alexanderpsergeev@gmail.com*

*b) datarasov@yandex.ru*

*c) bagalex3@gmail.com*

*d) and@ecko.uran.ru*

*e) adi8@yandex.ru*

*f) medvedev@ecko.uran.ru*

**Abstract** Modeling of spatial distribution of pollutants in the urbanized territories is difficult, especially if there are multiple emission sources. When monitoring such territories, it is often impossible to arrange the necessary detailed sampling. Because of this, the usual methods of analysis and forecasting based on geostatistics are often less effective. Approaches based on artificial neural networks (ANNs) demonstrate the best results under these circumstances. This study compares two models based on ANNs, which are multilayer perceptron (MLP) and generalized regression neural networks (GRNNs) with the base geostatistical method - kriging. Models of the spatial dust distribution in the snow cover around the existing copper quarry and in the area of emissions of a nickel factory were created. To assess the effectiveness of the models three indices were used: the mean absolute error (MAE), the root-mean-square error (RMSE), and the relative root-mean-square error (RRMSE). Taking into account all indices the model of GRNN proved to be the most accurate which included coordinates of the sampling points and the distance to the likely emission source as input parameters for the modeling. Maps of spatial dust distribution in the snow cover were created in the study area. It has been shown that the models based on ANNs were more accurate than the kriging, particularly in the context of a limited data set.

**Keywords:** Artificial Neural Networks, Kriging, Spatial Distribution, Modeling, Snow Cover

## 1. INTRODUCTION

In recent years, models based on ANNs have become a frequent practice. In (Abu Al Foul, 2012) ANNs were applied to predict the demand for electricity in Jordan, using the gross domestic product, population, exports and imports as four independent variables. The model, built based on ANNs, gave quite accurate energy demand forecast in this territory. In the study (Ogwueleka F.N. et. al., 2014) the model was built using ANNs to analyze and predict road traffic accidents. The population base, the number of road vehicles and the number of accidents were used as variables. Evaluation of the effectiveness identified the advantage of the neural network over other statistical methods. A number of studies were based on the application of hybrid models including ANNs and on the comparison of ANNs with other approaches. So, in (Anyama et. al., 2015) the model based on ANNs in conjunction with linear regression method was used to predict the outcome of a football match. This technique helped to achieve the prediction accuracy of 90.32%, which is significantly higher than when using other existing methods.

An increasing number of researchers apply ANNs in ecology, including in areas where geostatistical methods were previously used. Thus, in (Wahid H. et. al., 2013) ANNs were applied to estimate the concentration of ozone in the air basin of Sydney, Australia. The model provides more reliable assessment results and offers more accurate

predictions of ozone concentration. ANNs have been widely adapted and applied in practice by researchers (Liu Ze Lin et. al., 2010) in the light of increasing concerns about environmental problems such as global warming, frequent El Niño phenomena and atmospheric circulation anomalies. Methodology, based on ANNs, was applied to environmental planning, modeling and creation of high-quality digital maps of soil in the Rhineland-Palatinate (Germany), with a surface area of about 600 km<sup>2</sup> (Behrens Thorsten et. al., 2005). The authors show that this approach is cost effective and provides reliable results. ANNs were used to predict long-term changes of soil properties and development of such processes as degradation or desertification, which are one of the most important challenges of remote sensing. In (Kyung Hwa Cho et. al., 2011) the prediction effectiveness of four different models was evaluated; in particular, multiple linear regression (MLR), principal component regression (PCR), artificial neural network (ANN), and a combination of the principal component regression and the artificial neural network (PC-ANN) were applied to predict and to create a tool to assess the concentration of arsenic in Southeast Asia, including Cambodia, Laos and Thailand. Modeling results show that the hybrid model PC-ANN has demonstrated the best prediction accuracy among the four different models. The study (Anagu I. et. al., 2009) was conducted to develop models of sorption depending on basic properties of soils using ANNs. In the study, the data on soils collected on 133 agricultural lands throughout Germany was used. The results were compared with the data obtained by using the multiple linear regression (MLR). Models characteristics were assessed based on mean-root-square error, average error and modeling efficiency. It has been shown that the ANNs efficiency factors were in generally better than the MLR efficiency factors. In (Sarmadian F. et. al., 2008) performance of neural network model and multiple linear regressions were evaluated using a set of test data. The results showed that the artificial neural network with two neurons in the hidden layer demonstrated high performance in predicting soil properties. In (Sergeev A.P. et. al., 2015) also shows the advantage of models based on neural networks over geostatistical approaches.

In geostatistics a method of kriging is widely used. The efficiency of applying the kriging as an interpolation method depends on the actual spatial irregularity in the distribution of modeled pollutants. In the context of heterogeneous environment, this makes use of this interpolation method ineffective. Ordinary kriging is characterized by undervaluation in locations with high values and overvaluation in locations with low values.

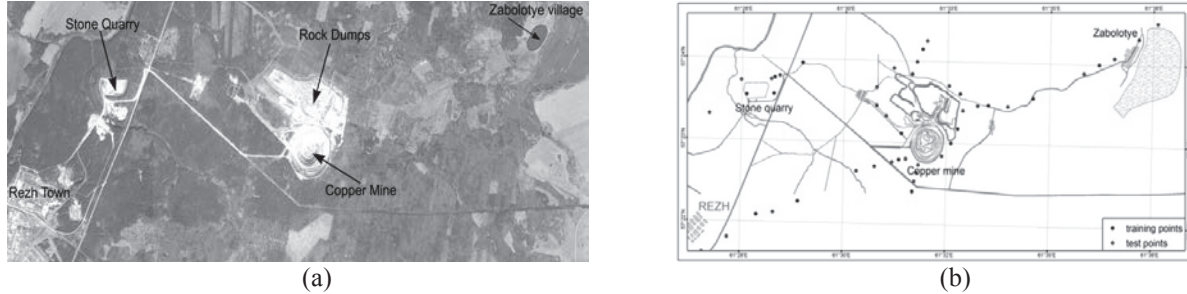
In our study two approaches are compared: modeling by the ANNs method and geostatistical analysis (kriging) to evaluate and predict the spatial dust distribution in the upper layer of the snow cover in the territory where there is a number of existing industrial facilities: copper quarry, stone quarry and Rezhevskoy nickel factory. Various natural media are monitored in this territory to assess the impact of industrial objects. An effective way of assessing the current contamination is snow sampling, bearing information on the number and composition of the pollutants emitted during the time of the snow cover build-up (Baglaeva E.M. et. al., 2012). Subsequent to the results of the snow sampling, the main indicator in evaluating the atmospheric pollution level in the studied territory compared with the background concentration is the indicator of "suspended solids" (dust). It is an integral indicator, which characterizes the general man-caused impact on aerial environment of the studied territory. However, snow sampling is a tedious process that does not allow carrying it out in sufficient detail. Thus, when assessing the pollution level only a limited data set can be used. And given that modeling and forecasting of real processes occurring in the atmosphere is a difficult task due to their non-linearity and high spatial heterogeneity, new approaches must be used.

## 2. MATERIALS AND METHODS

The snow sampling carried out during the period of maximum water content in snow in March of the year 2015. 47 samples were selected in total. Sampling points are situated on the investigated area around the exiting quarry of the Saf'yanovskiy copper-sulphide ore deposits and along the line, connecting Rezhevskoy nickel factory, the center of the copper quarry and the village Zabolotie, situated according to the wind rose prevailing in winter (Figure 1).

The actual point location was determined when sampling directly in-situ, bearing in mind the need for sampling in areas with intact (based on visual characteristics) snow cover. Snow samples were selected in accordance with RD 52.04. 186-89 "Guidelines on the control of atmospheric pollution". The entire vertical cut of the snow cover except for the lower 2-3 cm (to avoid contamination of samples with soil particles) was subject to the analysis. The samples were selected by the method of a square envelope with a side of 2 m (4 core-samples in the corners of the envelope and 1 core-sample in the middle). Envelope dimensions changed depending on the size of the corresponding area with the intact snow cover. The sufficient sample mass needed to perform the chemical analysis (not less than 3 kg) was achieved by adding snow cores sampled inside the envelope. The taken samples were placed

in double plastic bags, each of which was marked with a sample number in accordance with the sampling scheme, and recorded in the field log indicating the situational specifics of a sampling point location. To ensure the integrity of samples from melting (with the possible increase in air temperature), they were stacked densely in boxes made of corrugated cardboard. After that, the boxes were sealed, labelled and transferred to the control laboratory of JSC "Saf'yanovskaya Med'" for the chemical analysis. The values of dust concentrations in snow samples were used for modeling.



**FIGURE 1.** (a) Snow sampling area (background-Google Earth 7.1.5.1557). (b) the Scheme of the snow sampling points location

The coordinates of the sampling points and the distance to the likely emission source were used as input parameters for the models. Further, the models were designated as follows: abbreviation ANNs, with the input parameters of the model placed inside the parentheses. For example, MLP (x, y) denotes a model based on the multilayer perceptron with a direct sequence spread signal and a training Levenberg-Marquardt method, input parameters of which are the coordinates of the sampling point; and GRNN (x, y, dist) is a model based on generalized regression neural network, input parameters of which are the coordinates and distance to the likely emission source such as the pipe of the Rezhevskoy Nickel factory.

There are several types of kriging, including simple, ordinary, universal, indicator kriging and many others. The most applicable is the ordinary kriging. In the ordinary kriging an average value is considered to be permanent, but it is unknown. In addition, when using the local estimation the ordinary kriging does not require constancy of the average value throughout the whole assessment area; it is assumed that the average value is constant only at an estimation point. Number of data used in the assessment, and the values of weighting factors may vary depending on the location of the estimated point. Data is selected in a certain area around the point of estimation. The size and shape of this area depend on the source data. Thus, it is proposed to use the area oriented by the correlation ellipse. Reducing the area allows obtaining highly variable (less smoothed) estimate. Weights of ordinary kriging are obtained from the kriging equations, using a semivariogram. An unbiased estimate of the semivariogram function is a half of the root-mean-square deviation between the pair values. The semivariogram characterizes the degree of differences between data depending on the distance between them. The closer the data values (less the difference between them), the more the value of the semivariogram (1) (Matheron G., 1963):

$$\gamma(h) = \frac{\sum_{i=1}^{N(h)} |z(x_i) - z(x_i+h)|^2}{2N(h)} \quad (1)$$

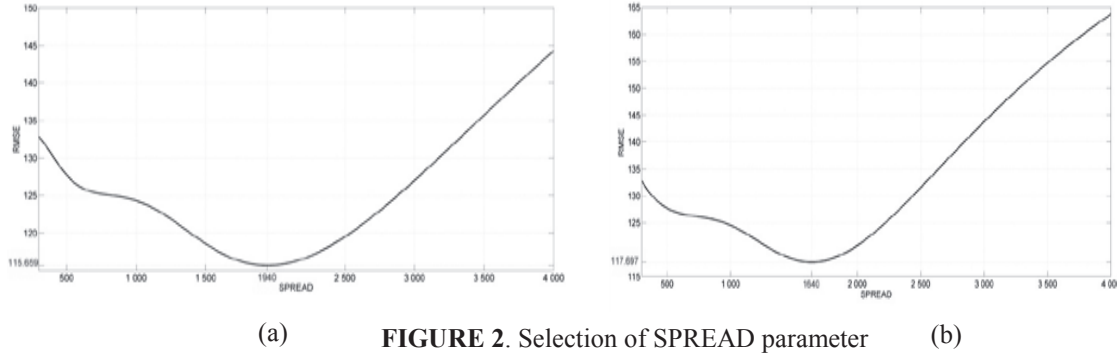
where  $\gamma(h)$  is the value of a semivariogram at distance interval  $h$ ;  $N(h)$  is the number of sample pairs at distance interval  $h$ ;  $z(x_i)$  and  $z(x_i+h)$  are the values for two points separated by a distance  $h$ . The variogram characterizes the differences of data points depending on the distance between them. The accuracy of kriging depends on the density and size of sampling spots. Therefore, a more efficient method is required to improve the accuracy of interpolation for producing high-resolution pollutants distribution maps.

The multilayer perceptron with a direct sequence spread signal and a training Levenberg-Marquardt method consists of an input layer, a hidden layer, and an output layer. The training rule is used to adjust the weights and offsets of the perceptron in order to approximate the output value to the target value. There is a large number of transfer functions such as: threshold, logistic, linear, hyperbolic tangent, restricted linear, radial basic functions and others. Typically, the transfer functions of all neurons in the network are fixed, and the weights are the network parameters that can vary. An error for the specific configuration of the network is determined by scanning of all available observations through the network and by comparing really given output values with the desired (target) values. All such differences are summarized in error function which value is a network error.

Generalized regression neural networks (GRNNs) are designed to meet the challenges of the generalized regression, time series analysis and function approximation. A characteristic feature of these networks is very high speed of their training (Voronov et. al., 2007).

ANN programming was conducted in the MATLAB® environment using the GUI interface. The multilayer perceptron was used. Its input layer consisted of sampling points coordinates, the hidden layer consisted of a few neurons and the output layer represented the dust concentration in the corresponding sample. The number of neurons in the hidden layer was selected by the least root-mean-square error of the dust concentration prediction. The number of neurons ranged from one to twenty. The quality of network training was checked by the correlation coefficient and the root-mean-square error between the result of the network running and the training data set. Multilayer perceptron with 7 neurons in the hidden layer provided the minimum value of the root-mean-square error for the training data set. Further research was conducted for a neural network with 7 neurons in the hidden layer.

To train the GRNN network and to predict the values in the test data set the MATLAB has been used. The *SPREAD* parameter characterizes the approximation properties of the GRNN network affects the accuracy of interpolation. If the influencing parameter *SPREAD* is low, then the radial basis function is characterized by a sharp recession and the input value range, to which neurons in the hidden layer respond, turns out to be small. With increasing *SPREAD* parameter, the slope of the radial basis function becomes smoother, and in this case, now a few neurons respond to the value of the input vector. Then at the network output the vector is created, corresponding to the average value of several target vectors, which correspond to the input vectors of the training data set, close to this input vector. The higher the value of the *SPREAD* parameter, the greater the number of neurons involved in the generation of the average value, and as a result the function generated by the network becomes smoother. The root-mean-square error between the predicted values in the testing data set and its actual values for each location were used for the selection of the *SPREAD* parameter. In this study, the *SPREAD* parameter value of 1940 provides the minimum root-mean-square error of interpolation (115.7) for the GRNN model (x, y) and the *SPREAD* parameter of 1640 provides the min root-mean-square error of interpolation (117.7) for the GRNN model (x, y, dist) (Figure 2).



(a) **FIGURE 2.** Selection of SPREAD parameter (b)

The dust concentration distributions in snow cover were built by means of the ordinary kriging based on chemical analysis data using ArcGIS software.

To compare the predictive ability of the methods by using the tool "Create a subset of data" ArcGIS Esri Geostatistical Analyst the sample was divided into two sub-samples, which are training and testing subsamples at the ratio of 50/50. The division was done in such a way that the odd points became the training subsample, and the even points became the testing subsample. The training sub-sample (23 points) was used as a training data set. Then the dust concentration values were predicted by kriging, MLP and GRNN using a testing sub-sample (24 points).

To assess the effectiveness of the different modelling methods three indices were used: the mean absolute error (MAE), the root-mean-square error (RMSE) and relative root-mean-square error (RRMSE), calculated as follows:

$$MAE = \frac{\sum_{i=1}^n |y_{Modi} - y_i|}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_{iMod} - x_i)^2}{n}} \quad (3)$$

$$RRMSE = \sqrt{\frac{\sum_{i=1}^n \frac{(x_{iMod} - x_i)^2}{x_i^2}}{n}} \quad (4)$$

where  $x_{iMod}$  is a predicted concentration (ANNs, kriging),  $x_i$  is a measured concentration,  $n$  is a number of points.



### 3. RESULTS AND DISCUSSION

Table 1 summarizes the main statistics of the dust concentration in the study area. Dust concentration was positively skewed and showed leptokurtic distribution. Table 2 shows the accuracy assessment indices of the estimated dust concentrations at validation sites for the different methods of modeling.

**TABLE 1.** Dust concentration, descriptive statistics

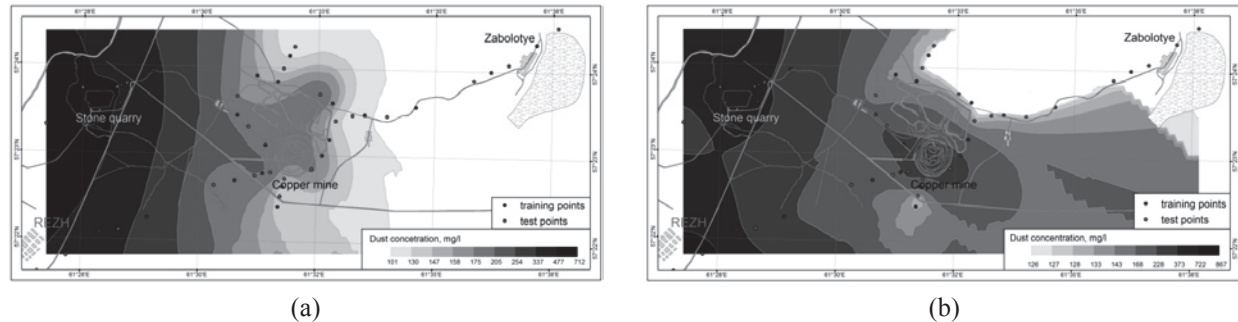
	Valid N	Mean	Median	Min.	Max.	Lower	Upper	Variance	Std.Dev.	Skewness	Kurtosis
Dust conc., mg/l	47	248,1	173,8	39,8	1011	90,7	375,1	48464	220,1	1,7	2,5

**TABLE 2.** The accuracy assessment indices of the estimated dust concentrations

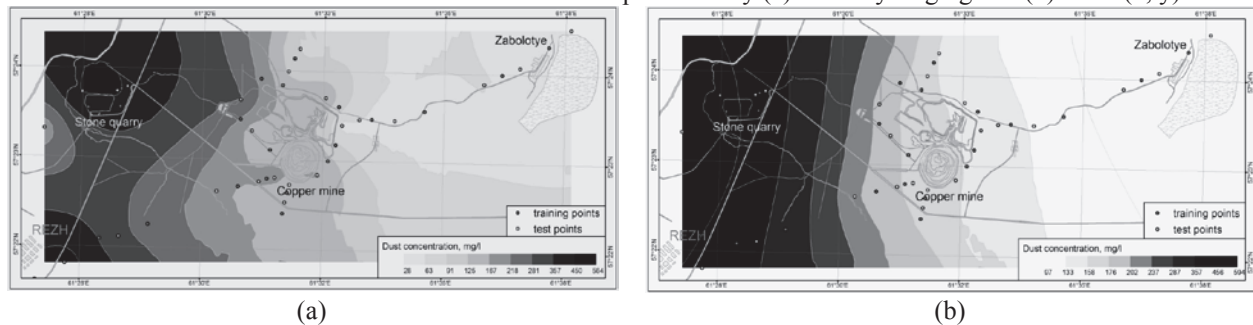
Model	RMSE	RRMSE	MAE
MLP (x, y)	86.2	0.72	72.8
MLP (x, y, dist)	117.6	0.49	86.7
GRNN (x, y)	117.7	0.50	81.1
GRNN (x, y, dist)	115.7	0.49	80.3
Kriging	124.6	0.49	81.8

Taking into account all indices, the model based on generalized regression neural network turned out to be the most accurate, the input parameters of which included coordinates of the sampling points and the distance to the likely emission source GRNN (x, y, dist). The indices RMSE and MAE for the multilayer perceptron-based model, at the entry point of which there were only MLP (x, y) coordinates, turned out to be significantly lower than for other models. At the same time, the RRMSE index was much higher. When adding the information about the distance to the likely emission source to the MLP model, the RRMSE index decreased significantly, and RMSE and MAE indices rose. For kriging, the RMSE index turned out to be significantly higher.

Figures 3–4 show the dust concentration distribution in the snow cover of the study area based on the assessments of the various models.



**FIGURE 3.** Distribution of dust concentration predicted by (a) ordinary kriging and (b) MLP (x, y)



**FIGURE 4.** Distribution of dust concentration predicted by (a) MLP (x, y, dist), (b) GRNN (x, y)

## 4. CONCLUSION

Models based on artificial neural networks turned out to be suitable for predicting the dust concentration distribution in the snow cover in a limited number of sampling points. They outdid the geostatistical methods in the prediction accuracy. Adding information about the distance to the likely emission source increased the accuracy of the model based on GRNN, the RRMSE index for the MLP model (x, y, dist) also proved to be significantly lower. However, the RMSE index for the MLP model (x, y, dist) on the contrary, increased. Most likely this is due to the fact that the MLP model (x, y) accurately predicted values at the points where the dust concentration in samples was relatively high, in the case of relatively low concentrations of dust the accuracy of the MLP model (x, y) prediction was lower. When adding new information about the distance to the likely emission source, the MLP model started predicting values more accurately at the points with relatively low dust concentrations and worse at the points with high dust concentration.

## REFERENCES

1. Abu Al Foul. (2012) Forecasting Energy Demand in Jordan. Topics in Middle Eastern and African Economies 14. 473-478.
2. Anagu I., Ingwersen J., Utermann J., Streck T. (2009) Estimation of heavy metal sorption in German soils using artificial neural networks. [Geoderma](#). 152 . 104–112.
3. Anyama, Oscar Uzoma, Igiri, Chinwe Peace. (2015) An Application of Linear Regression. International Journal of Engineering Research & Technology (IJERT) № 4 (01). 457-461.
4. Baglaeva E. M., Sergeev A. P., Medvedev A. N. (2012) Spatial structure of technogenic pollution of snow cover by soluble and insoluble forms of metals in the industrial city and its neighbourhood. Geoecology, engineering geology, hydrogeology, geocryology. 2012, Issue 4, 326-335 pp. (in Russian).
5. Behrens Thorsten, Forster Helga, Scholten Thomas, Steinrucken Ulrich, Spies Ernst-Dieter, Goldschmitt Michael. (2005) Digital soil mapping using artificial neural network. Journal of Plant Nutrition and Soil Science. No. 168. 1–13.
6. Kyung Hwa Cho, Sthiannopkao Suthipong, Pachepsky A Yakov., Kyoung-Woong Kim, Joon Ha Kim. (2011) Prediction of contamination potential. Water research. No. 45. 5535–5544.
7. Liu Ze Lin, Peng Chang Hui, Xiang Wen Hua, Tian Da Lun, Deng Xiang Wen, Zhao Mei Fang. (2010) Application of artificial neural networks. Chinese Science Bulletin. No.34. 3853–3863.
8. Matheron G. (1963) Principles of geostatistics, [Economic Geology](#) 58, 1246–1266.
9. Ogwueleka F.N., Sanjay Misra, Ogwueleka T.C., L. Fernandez-Sanz. (2014) An Artificial Neural Network Model for Road Accident Prediction. Acta Polytechnica Hungarica 11 (5).
10. Sarmadian F., Mehrjardi R. Taghizadeh. (2008) Modeling of Some Soil Properties Using Artificial Neural Network. Global Journal of Environmental Research. No 1. 30–35.
11. Sergeev A.P., Buevich A.G., Medvedev A., Subbotina I.E., Sergeeva M. (2015) Artificial neural network and kriging interpolation for the chemical elements contents in the surface layer of soil on a background area. 15th International Multidisciplinary Scientific GeoConference SGEM 2015, Conference Proceedings, 2015, Book 3 Vol. 2. 49–56.
12. Voronov I.V., Politov E.A., Efremenko V.M. (2007) Overview of types of artificial neural networks and their teaching methods. Vestnik KuzGTU. № 3. 38-42. (in Russian).
13. Wahid H., Ha Q.P., Duc H., Azzi M. (2013) Neural network-based meta-modelling approach. [Applied Soft Computing](#). No 13. 4087–4096.